Chromosomal profiles of high-grade cervical intraepithelial neoplasia relate to duration of preceding high-risk human papillomavirus infection
Chromosomal profiles of CIN3 relate to duration of hrHPV infection

Abstract

High-grade cervical intraepithelial neoplasia (CIN2/3) represents a heterogeneous disease both with respect to clinical behaviour and chromosomal aberrations detected. We hypothesised that the extent of chromosomal aberrations reflects the duration of their existence.

Chromosomal profiles were determined of CIN3 of women with a known 5 year history of high-risk human papillomavirus virus (hrHPV) infection, in which duration of prior hrHPV infection was considered a proxy for duration of CIN3 existence. Eleven women had a <5 year preceding hrHPV infection (CIN3<5yrPHI) and 24 had a PHI lasting ≥5 years (CIN3≥5yrPHI). For comparison, 6 CIN3 adjacent to squamous cell carcinomas (CIN3-SCC), the corresponding SCC and 6 CIN1 were included.

Unsupervised hierarchical clustering analysis of the chromosomal profiles revealed two clusters. One was characterised by a low number of chromosomal aberrations and included all CIN1, 81.8% of CIN3<5yrPHI and 33.3% of CIN3≥5yrPHI. Samples in the second cluster, displaying multiple aberrations, included 18.2% of CIN3<5yrPHI, 66.7% CIN3≥5yrPHI, all except one CIN3-SCC and all SCC. The number of genomic aberrations increased according to lesion grade and also with longer duration of PHI. The increase in aberrations in CIN3≥5yrPHI compared to <5yrPHI was highly significant (p=0.001), suggesting that CIN3≥5yrPHI represent more severe lesions.

In conclusion, longer duration of preceding hrHPV infection is associated with an increased number of chromosomal aberrations. Hence, CIN3 with a longer duration of existence are likely more prone to have an increased short-term risk of cervical cancer.
Introduction

Infection with high-risk human papillomaviruses (hrHPVs) has been causally related to cervical cancer development. More than 80% of hrHPV infections are cleared by the immune system. In some instances precancerous lesions arise, so-called cervical intraepithelial neoplasia (CIN). Histologically, CIN can be divided into low-grade (CIN1, mild dysplasia) and high-grade (CIN2/3, moderate/severe dysplasia). Not all CIN represent true precursor stages of cervical cancer. Low-grade CIN particularly represent productive infections. High-grade CIN, prone to progress to invasive carcinomas, is associated with transforming infections, characterised by deregulated expression of viral oncoproteins E6 and E7 in proliferating cells. The E6 protein mediates p53 degradation, thereby preventing cell cycle arrest and apoptosis upon DNA damage. E7 mediates inactivation of pRb family members, impairing G1/S cell cycle control and contributing to uncontrolled proliferation. Deregulated E6/E7 activity in cycling cells results in increased expression of p16INK4a, which is considered a marker for CIN harbouring transforming hrHPV infections. Together, E6 and E7 expression in dividing cells leads to accumulation of specific (epi)genetic changes in the host cell genome, likely driving malignant progression.

High-grade CIN may be rapidly induced within 2-3 years following hrHPV infection, whereas progression to invasive carcinoma may take another 10-30 years. Most newly acquired hrHPV infections, particularly those in older women (≥43 years), only rarely give rise to high-grade CIN within a period of 3 years, while long-term hrHPV infections within this age category confer substantial risk of progression to CIN3, suggesting these types of infections represent different stages of hrHPV natural history. High-grade CIN represents a heterogeneous disease with only advanced stages likely to have short-term invasive potential. Using comparative genomic hybridisation microarray (arrayCGH) we previously found that within a group of histologically similar, p16INK4a-positive high-grade CIN, one subset showed few chromosomal aberrations while the other showed a substantial number of aberrations resembling invasive cervical carcinomas, suggesting the latter subset are longer existing, more advanced lesions with high short-term progression risk. In current clinical practice high-grade CIN are radically treated to prevent progression to cervical cancer, thus, no natural history studies can be performed to corroborate this hypothesis.

Since continuous E6/E7 expression is thought to drive accumulation of chromosomal aberrations underlying progressive disease it is likely that the degree of chromosomal abnormality in CIN depends on duration of hrHPV infection. In this context, valuable information can be gathered from chromosomal profiling studies of high-grade CIN associated with short-term versus long-term hrHPV infections. Here, we performed such a study using information of stored cervical samples collected during the course of the longitudinal, population-based screening trial POBASCAM. Chromosomal profiles were determined of CIN3 diagnosed in the second screening round of women who had hrHPV infections at baseline, 5 years
earlier (long-term infection), versus CIN3 of women who were hrHPV-negative at baseline and acquired hrHPV infections after study entrance (short-term infection).

**Materials and Methods**

**Tissue specimens**
From 35 women participating in POBASCAM\(^9\) (trial# ISRCTN20781131) formalin-fixed paraffin-embedded (FFPE) CIN3 biopsy material was used. Women had normal cytology at study entrance and without any intervention a CIN3 was detected in the subsequent screening round, approximately 5 years later. In 11 women no hrHPV infection was detected in the baseline sample by GP5+/6+ analysis (method described in\(^{10}\)). These women acquired infections after study entrance and were classified as having CIN3 with <5 year preceding hrHPV infection (CIN3<5yrPHI, short-term infection). In 24 women hrHPV was detected at baseline already. These women were classified as having CIN3 with ≥5 year preceding hrHPV infection (CIN3≥5yrPHI, long-term infection). For CIN3≥5yrPHI it was ensured the hrHPV type in the biopsy matched that detected at baseline. None of the women had a history of (pre)malignant anogenital disease before study enrolment.

FFPE biopsies of 6 CIN1 were included from women participating in POBASCAM. All had normal cytology at baseline and were diagnosed with an hrHPV-positive CIN1 in the next screening round. Four were hrHPV-negative and 2 hrHPV-positive at baseline, of which one harboured the same hrHPV type in the CIN1 and baseline sample. Six CIN3 adjacent to SCC, considered representatives of advanced disease, and their SCC counterparts were collected at the Departments of Obstetrics and Gynaecology (VU University Medical Center, Amsterdam, the Netherlands) during routine clinical practice. The average age of women was 41 years (range: 34-61), with no significant difference between women diagnosed with CIN3<5yrPHI and CIN3≥5yrPHI. Histological review of lesions was performed by experienced pathologists (F.J.v.K, M.C.G.B). To ensure CIN3 and SCC harboured transforming hrHPV infections, immunohistochemical staining for p16\textsuperscript{INK4a} was performed as described previously.\(^8\) All CIN3 and SCC showed diffuse p16\textsuperscript{INK4a} staining.

**ArrayCGH of microdissected tissues**
Amplified DNA, from microdissected dysplastic areas of tissue specimens, was hybridised to 105K arrays (Agilent Technologies, Palo Alto, USA) using a pool of 5 normal cervical tissues as reference (see Supporting File I for sample preparation and procedures). Array data is available from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/projects/geo/) through series accession number GSE30155.
Hierarchical clustering

Unsupervised hierarchical clustering of low- and high-grade CIN plus SCC was performed using a modified version of Weighted Clustering of Called aCGH data (WECCA_v._0.3).\textsuperscript{11,12} In the dendrogram construction Ward’s linkage was applied to obtain compact and well-separated clusters.

Statistical analysis

The percentage of oligonucleotides deviating from the normal state as determined by the CGHcall algorithm (Supporting File I) between the various groups of CIN and SCC was compared using the non-parametric Mann-Whitney U test. GAORT (Global Copy Number Aberration Odds Ratio Test) was used to test whether the odds of an aberrant region in one sample group was systematically larger than that of another group (Supporting File II).

Results

Distinct chromosomal profiles of CIN3 with <5yrPHI and ≥5yrPHI

Following microdissection, DNA from 11 CIN3<5yrPHI, 24 CIN3≥5yrPHI, 6 CIN3 adjacent to SCC and their corresponding SCC and 6 CIN1 were subjected to high-resolution arrayCGH. To determine in an unbiased manner whether duration of preceding hrHPV infection was associated with the chromosomal profiles of CIN3, unsupervised hierarchical clustering analysis was performed. Two main clusters were identified (Figure 1). Cluster 1 contained 24 samples with few chromosomal aberrations while cluster 2 contained 29 samples with multiple aberrations. Cluster 1 included all CIN1, 9/11 CIN3<5yrPHI (81.8%), 8/24 CIN3≥5yrPHI (33.3%) and 1/6 CIN3 adjacent to SCC (16.7%). Cluster 2 contained 2/11 CIN3<5yrPHI (18.2%), 16/24 CIN3≥5yrPHI (66.7%), 5/6 CIN3 adjacent to SCC (83.3%) and all SCC. Distribution of CIN3<5yrPHI and CIN3≥yrPHI over clusters 1 and 2 was significant (p=0.012), as determined with the Fisher exact test. Four of the 6 CIN3 adjacent to SCC clustered together with their SCC counterpart, indicating great similarity between chromosomal profiles.

The majority of CIN3≥5yrPHI exhibited chromosomal profiles resembling those of CIN3 adjacent to SCC and corresponding SCC whereas the majority of CIN3<5yrPHI did not.

The major difference in chromosomal aberrations occurring in CIN3 of clusters 1 and 2 was determined by the maximum pair-wise symmetrised Kullback-Leibler divergence (Supporting Figure I). Gain of chromosome 3 for CIN3 in cluster 2 appeared one of the most striking differences.
Chromosomal profiles of CIN3 relate to duration of hrHPV infection

Figure 1: Results of unsupervised hierarchical clustering of different types of CIN and SCC. Light boxes: CIN3 with <5yrPHI; darker boxes: CIN3 with ≥5yrPHI. SCC:CIN3: CIN3 adjacent to SCC; SCC: carcinoma part of the SCC. CIN1: CIN1. Gains are represented in green, losses in red and amplifications in white. Chromosomes run from bottom to top, starting at the bottom is the p-arm, followed by the q-arm.
The number of chromosomal aberrations increases with a longer duration of preceding hrHPV infection

The percentage of oligonucleotides deviating from the normal state increased per lesion grade from 0.16% in CIN1, to 2.83% in CIN3<5yrPHI, 16.52% in CIN3≥5yrPHI, 28.75% in CIN3 adjacent to SCC and 34.27% in SCC. Similar trends were seen for losses and gains separately (Table 1).

CIN1 had significantly fewer chromosomal losses than CIN3<5yrPHI and CIN3≥5yrPHI (p=0.035 and p=0.015, respectively; Table 2). Although the number of gains was increased in both groups of CIN3 compared to CIN1, this was only significant for CIN3≥5yrPHI (p<0.001). Interestingly, CIN3≥5yrPHI had significantly more aberrations than CIN3<5yrPHI (p=0.001), particularly due to an increased number of gains in this group (p=0.001). The increase in lost chromosomal regions in CIN3≥5yrPHI was not significant (p=0.303). CIN3 adjacent to SCC and their SCC counterparts had significantly more losses compared to both CIN3<5yrPHI (p=0.016 and p=0.001, respectively) and CIN3≥5yrPHI (p=0.017 and p=0.002, respectively). While the number of gains in CIN3 adjacent to SCC and their SCC counterparts was significantly increased compared to CIN3<5yrPHI (p=0.035 and p=0.001 respectively), this was not the case for CIN3≥5yrPHI.

Global Copy Number Aberration Odds Ratio Test (GAORT) was used to determine whether CIN3≥5yrPHI showed a systematic increase in the odds of a certain aberrated region compared to CIN3<5yrPHI. This test supposes that if CIN3≥5yrPHI develop from CIN3<5yrPHI, aberrations found in CIN3≥5yrPHI include at least aberrations found in CIN3<5yrPHI. An increase in odds of an occurring aberration corroborates the idea that CIN3≥5yrPHI originate from CIN3<5yrPHI. There was a significant systematic increase in chromosomal aberrations in CIN3≥5yrPHI compared to CIN3<5yrPHI (p<0.001). In particular the odds of gained regions were significantly increased (p<0.001). Hence, it may be concluded that CIN3≥5yrPHI are further progressed than CIN3<5yrPHI.

Table 1: The percentage of array oligonucleotides deviating from the normal state in the different groups of CIN and SCC.

<table>
<thead>
<tr>
<th></th>
<th>% LOSSES</th>
<th>% GAINS</th>
<th>% TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>average</td>
<td>range</td>
<td>average</td>
</tr>
<tr>
<td>CIN1</td>
<td>0.07%</td>
<td>0.01-0.12%</td>
<td>0.09%</td>
</tr>
<tr>
<td>CIN3&lt;5yrPHI</td>
<td>0.95%</td>
<td>0-3.06%</td>
<td>1.87%</td>
</tr>
<tr>
<td>CIN3≥5yrPHI</td>
<td>3.77%</td>
<td>0.02-30.70%</td>
<td>12.74%</td>
</tr>
<tr>
<td>SCC-CIN3</td>
<td>15.12%</td>
<td>0.07-37.40%</td>
<td>13.63%</td>
</tr>
<tr>
<td>SCC</td>
<td>18.43%</td>
<td>4.78-34.67%</td>
<td>15.84%</td>
</tr>
</tbody>
</table>
Chromosomal profiles of CIN3 relate to duration of hrHPV infection

Table 2: The percentage of array oligonucleotides deviating from the normal state in the different groups of CIN and SCC.

<table>
<thead>
<tr>
<th></th>
<th>CIN3&lt;5yrPHI</th>
<th>CIN3≥5yrPHI</th>
<th>CIN3-SCC</th>
<th>SCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOSSES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIN1</td>
<td>0.035</td>
<td>0.015</td>
<td>0.016</td>
<td>0.004</td>
</tr>
<tr>
<td>CIN3&lt;5yrPHI</td>
<td>0.303</td>
<td>0.016</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>CIN3≥5yrPHI</td>
<td>0.017</td>
<td></td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>CIN3-SCC</td>
<td></td>
<td></td>
<td></td>
<td>0.873</td>
</tr>
<tr>
<td>GAINS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIN1</td>
<td>0.119</td>
<td>3.47×10^-4</td>
<td>0.0037</td>
<td>0.004</td>
</tr>
<tr>
<td>CIN3&lt;5yrPHI</td>
<td>0.001</td>
<td>0.0035</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>CIN3≥5yrPHI</td>
<td></td>
<td>0.667</td>
<td>0.233</td>
<td></td>
</tr>
<tr>
<td>CIN3-SCC</td>
<td></td>
<td></td>
<td>0.631</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIN1</td>
<td>0.016</td>
<td>1.89×10^-4</td>
<td>0.0037</td>
<td>0.004</td>
</tr>
<tr>
<td>CIN3&lt;5yrPHI</td>
<td>0.001</td>
<td></td>
<td>0.0021</td>
<td>0.001</td>
</tr>
<tr>
<td>CIN3≥5yrPHI</td>
<td></td>
<td>0.162</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td>CIN3-SCC</td>
<td></td>
<td></td>
<td>0.522</td>
<td></td>
</tr>
</tbody>
</table>

Frequently altered regions in CIN3<5yrPHI and CIN3≥5yrPHI

The frequency of gains and losses per oligonucleotide was plotted for CIN1, the different groups of CIN3 and SCC to determine common and specific chromosomal aberrations (Figure 2A). The frequency plots confirmed that duration of hrHPV infection and number of aberrations increased in parallel. Interestingly, the SCC components in general displayed the same aberrations as their adjacent CIN3, in line with clustering results (Figure 1). Extra aberrations in these CIN3 and SCC compared to CIN3≥5yrPHI included gain of 5p and loss of 11q, each occurring in ≥50.0% of these samples. Additional losses (e.g. at 2, 3p, 4, 7, 10p, 14 and 17) tended to be more frequently present in SCC compared to their adjacent CIN3.

Smallest regions of overlap (SRO) were determined in CIN3 with < and ≥5yrPHI and SCC to determine regions that may be essential for progression towards malignancy (Supporting Table I). Attention was directed towards aberrations occurring with a frequency of ≥25.0% in these combined groups, showing successive increases in frequency from CIN3<5yrPHI to CIN3≥5yrPHI and SCC with at least an incidence of 50% in the SCC. Gained regions were located on chromosomes 1p, 1q, 3q and 20q and a lost region on chromosome 2q. None of these regions were affected in CIN1. Gains of 3q occurred frequently in CIN3<5yrPHI, followed by regions on 1p and 1q. Gains at 20q and loss at 2q occurred with lesser frequency or were not present in CIN3<5yrPHI, suggesting that these aberrations occur later in the carcinogenic process following gains of 1p, 1q and 3q. ArrayCGH results were confirmed by FISH analysis (Supporting File III, Table II; and Figure II). A concept of cervical carcinogenesis in terms of the sequential acquisition of specific chromosomal aberrations based on present findings is shown in Figure 2B.
Figure 2: (A) Frequency plot of aberrations occurring in CIN1, the different groups of CIN3, and SCC. The percentages of gains (positive Y-axis) and losses (negative Y-axis) for each oligonucleotide are shown. (B) Concept of the sequential order of chromosomal aberrations occurring during cervical cancer development. The number of chromosomal aberrations increases with lesion grade and duration of preceding hrHPV infection. CIN3 related to a longer hrHPV exposure progressively acquire non-random gains at 3q, 1p, 1q followed by 20 and loss of 2q. Progression to invasive cancer is associated with additional aberrations; e.g. gain of 5p and loss of 11q.
Chromosomal profiles of CIN3 relate to duration of hrHPV infection

Discussion

To the best of our knowledge, this is the first study showing an association between number of chromosomal aberrations in CIN3 and duration of hrHPV infection. Particularly, CIN3 found in women with long-term preceding hrHPV infections (≥5 year) were characterised by a higher number of chromosomal aberrations.

Clustering analysis revealed that genomic profiles of a major subset (66.7%) of CIN3 with a long-term preceding hrHPV infection were more similar to invasive carcinomas, indicative of more advanced lesions. The percentage of genomic aberrations gradually increased per lesion grade from CIN1, via CIN3<5yrPHI, CIN3 ≥5yrPHI and CIN3 adjacent to SCC to SCC. This is in concordance with previous cross-sectional studies showing an increase in the number of aberrations from CIN1, to high-grade CIN and cancer. However, no relationship with duration of existence could be concluded from these studies.

Duration of hrHPV infection as proxy of lesion age not only enables estimation of the order of chromosomal events in cervical carcinogenesis in vivo, but also provides an explanation for the heterogeneity described in previous studies. Our data suggest that the first aberrations in high-grade CIN are losses, as the number of chromosomal losses is significantly increased in CIN3<5yrPHI compared to CIN1. CIN3≥5yrPHI showed significantly more gains than CIN3<5yrPHI, particularly at regions on chromosomes 1p, 1q and 3q, indicating these represent early events in progression of CIN3 to invasive cancer. CIN3 adjacent to SCC and the SCC counterparts showed an increased number of losses compared to CIN3≥5yrPHI, indicating this type of aberration to occur later during progression.

The smallest regions of overlap (SRO) revealed gained regions at 3q, 1p, 1q and 20q and a lost region at 2q to be frequent aberrations (≥25.0% occurrence in all groups). Whereas gains at 3q appeared to represent the earliest event, followed by gains at 1p and 1q, gains at 20q and loss at 2q represented later events in progression of CIN3, as these were primarily present in CIN3≥5yrPHI and SCC. Interestingly, SROs of frequently gained regions (i.e. 1p, 1q and 3q) overlap with those found in our previous cross-sectional study on 46 p16INK4a-immunopositive CIN2/3 analysed by BAC arrays. Most of these aberrations were also found by others using different platforms.

Interestingly, some SROs harbour genes previously found to be overexpressed in cervical carcinomas as a result of copy number gains, such as DTX3L and MCM2, both located within the SRO at 3q.

One of the study limitations is the relatively low number of CIN3<5yrPHI available. Of women participating in POBASCAM who had normal cytology at baseline, relatively few with newly acquired hrHPV infections developed CIN3 within 5 years. This is in agreement with the known low 3- and 5-year CIN3+ risk of hrHPV-negative women with normal cytology. However, the fact that even in this relatively small sample set clearly significant results were obtained indicates these findings are highly relevant and consistent. A second limitation is related to
the fact that POBASCAM could only provide hrHPV information dating up to 5 years prior to detection of the lesion and that neither the exact moment of infection nor the actual moment at which the lesion arose is known. Consequently, it can be speculated that the 33.3% of CIN3≥5yrPHI with few chromosomal aberrations may be related to a relatively short duration of hrHPV infection within this group. Conversely, the 2 CIN3<5yrPHI mapping to cluster 2, although having fewer aberrations than the other lesions in cluster 2, may have a PHI approaching 5 years. The finding that 1 CIN3 adjacent to SCC mapped to cluster 1 is exceptional and may be explained by heterogeneity within the lesion.

Present findings indicate that surrogate markers for long-term viral persistence, a well known risk factor for cervical cancer reflected by abnormal chromosomal profiles, would allow distinction between high-grade CIN with a short-term progression risk in need of immediate treatment (cluster 2 profile) and those that could be managed by close surveillance (cluster 1 profile). The latter may be particularly beneficial to women of reproductive age, as treatment of high-grade CIN coincides with some degree of morbidity of the cervix and can give rise to preterm delivery.21 Our data are particularly relevant for cervical cancer prevention strategies using primary hrHPV testing. We demonstrated that CIN3 in hrHPV-test negative women within a 5 year follow-up period typically have a low number of chromosomal aberrations and a likely low short-term progression risk. This is fully in line with the low carcinoma incidence in the second screening round of women who had been screened by hrHPV testing.22

Finally, we speculate that molecular markers, particularly markers specific for gene methylation events, which are associated with high-grade CIN displaying many chromosomal aberrations8, may be ideal biomarkers to triage hrHPV-positive women for colposcopic examination.
Chromosomal profiles of CIN3 relate to duration of hrHPV infection

Reference List


Chromosomal profiles of CIN3 relate to duration of hrHPV infection


Supporting File I: Sample preparation & arrayCGH procedures.

Microdissection, DNA extraction and amplification
The dysplastic areas of the tissue specimens were marked on 4 μm haematoxylin and eosin stained sections. Subsequently, serial 8 μm sections were cut and mounted on 2.0 μm PEN-foil membrane slides (Leica, Heidelberg, Germany) for laser capture microdissection. After deparaffinisation and haematoxylin staining, laser capture microdissection was performed using a Leica ASLMD microscope, to enrich for dysplastic cells. Dissected material was incubated overnight at 37°C with 1 M sodium thiocyanate to reduce the number of formalin-induced cross links between the DNA strands. After removal of sodium thiocyanate and washing with PBS, the samples were treated with 1 mg/ml proteinase K for 5 days with daily enzyme additions followed by DNA extraction using the Qiagen DNA micro kit (Qiagen, Westburg, Leusden, The Netherlands) according to the manufacturer’s protocol. To assess the DNA quality and in order to obtain sufficient DNA for microarray experiments, whole genome amplification using the Bioscore kit (Enzo Bioscore™ Screening and Amplification, Enzo Life Sciences, Farmingdale, USA) was performed as described by Buffart et al. The linearity of the amplification was validated by comparing the genomic profiles of the same sample with and without amplification. A normal reference pool was generated of formalin-fixed paraffin-embedded cervical biopsies of 5 women without malignant cervical disease who were negative for the presence of either low- or high-risk HPV. DNA was isolated using proteinase K as described above. This pool was first tested against a human blood reference derived from healthy women to ensure that there were no genomic aberrations present.

ArrayCGH procedures
Labeling of DNA with Cy3 or Cy5 was performed with the Enzo Genomic DNA labeling kit according to the manufacturer’s instructions (Enzo Life Sciences) using 500 ng of amplified DNA. Hybridisations were performed on slides containing two 105K arrays, with each array containing 99,000 synthetic 60-mer oligonucleotides (Agilent Technologies, Palo Alto, USA). Across array CGH (aaCGH) as well as across slide comparison was used, in which a sample was either hybridised to a reference, or to another sample as described by Buffart et al. To overcome potential wave bias due to differences in GC-content of the different chromosomal regions, a smoothing algorithm NoWaves version 0.3 was applied to the dataset. Gains and losses were determined using the statistical package CGHcall version 2.5.0. CGHregions version 2.3 was used to reduce the dataset to chromosomal regions, allowing for 0.1% information loss (threshold 0.001). Use of regions instead of single oligonucleotides has been shown to improve the effectiveness of subsequent statistical analyses and facilitates interpretation of the results.
Chromosomal profiles of CIN3 relate to duration of hrHPV infection

Supporting File II: Global Copy Number Aberration Odds Ratio Test (GAORT).

Global copy number Aberration Odds Ratio Test (GAORT) was used to test whether the odds of an aberrant region in one group of samples was systematically (all over the genome) larger than that of another group. A systematic increase in odds might be expected if both groups are precursors of the same cancer and the first group of samples can be viewed as being further progressed than the second group. GAORT models the aberration profiles by a logistic model, which incorporates an intercept for every genomic region (to allow for aberration signature differences between regions) and a regression coefficient shared by all regions reflecting the odds increase between the groups. The parameters of the model are estimated by means of likelihood maximization. The null hypothesis (of an odds ratio between the groups being 1) is evaluated by permutation of the group labels.

Let $Y_{ij} \in \{0,1\}$ indicates whether region $j = 1,\ldots,p$ of sample $i = 1,\ldots,n$ is non- or aberrated, respectively. Or equivalently, not lost or lost, et cetera. Furthermore, $X_i \in \{-1,1\}$ encodes for the group to which sample $i$ belongs. The CIN lesions with $X_i = 1$ have progressed further than the samples with $X_i = -1$. It is assumed that all samples progress along the same lines. The odds of an aberration may therefore be assumed to be higher in the former group than in the latter. Oversimplifying, we assume that the odds ratio between the groups is constant over the regions (proportional odds).

Model the probability of an aberration by:

$$P(Y_{ij} = 1) = \pi_j(X_i; \alpha_j, \beta) = \frac{\exp(\alpha_j + \beta X_i)}{1 + \exp(\alpha_j + \beta X_i)}$$

for all $i$ and $j$.

Thus, the $\beta$ is the same for all regions. Put differently, the odds ratio between the groups is identical for all regions. The log-likelihood is then given by:

$$\mathcal{L}(Y, X; \alpha_1, \ldots, \alpha_p, \beta) = \sum_{j=1}^{p} \sum_{i=1}^{n} Y_{ij} \log[\pi_j(X_i; \alpha_j, \beta)] + (1 - Y_{ij}) \log[1 - \pi_j(X_i; \alpha_j, \beta)],$$

with first order partial derivatives:

$$\frac{\partial \mathcal{L}}{\partial \alpha_j} = \sum_{i=1}^{n} Y_{ij} - \pi_j(X_i; \alpha_j, \beta)$$

and

$$\frac{\partial \mathcal{L}}{\partial \beta} = \sum_{i=1}^{n} X_i [Y_{ij} - \pi_j(X_i; \alpha_j, \beta)].$$

Estimate $\alpha_1, \ldots, \alpha_p, \beta$ by the following iterative procedure:

Step 1: Set $\hat{\beta}^{(0)} = 0$.

Step 2: Obtain $\hat{\alpha}_1^{(k)} \ldots, \hat{\alpha}_p^{(k)}$ by maximizing the likelihood with respect to the $\alpha$s keeping $\beta$ fixed at $\hat{\beta}^{(k-1)}$.

Step 3: Obtain $\hat{\beta}^{(k)}$ through likelihood maximization keeping the $\alpha$s fixed at their latest estimate.

Step 4: Iterate between step 2 and 3 until convergence.

To test the proportional odds assumption, we test $H_0: \beta = 0$ vs. $H_A: \beta > 0$. Note that $\beta = 0$ corresponds to an odds ratio of 1. Simply use the estimate of $\beta$ as a test statistic. The null distribution is obtained through permutation of the sample labels and re-estimation of $\beta$. 

48
Supporting File III: FISH on 3q and 8q validate arrayCGH results.

Materials & Methods

ArrayCGH results were validated using fluorescence in-situ hybridisation (FISH) on 3 μm sections of FFPE samples of 2 CIN3<5yrPHI and 7 CIN3≥5yrPHI using the cervical cancer kit with probes targeting TERC at 3q26 and C-MYC at 8q24 (Cervical Cancer, 3q26, 8q24, SE7, triple-colour, Kreatech diagnostics, Amsterdam, the Netherlands) according to the manufacturer’s protocol. The pepsin treatment was performed for 30 minutes. Samples were analysed with a Zeiss Axiophot fluorescence microscope (Carl Zeiss, Sliedrecht, Benelux) using CytoVision software (Applied Imaging Corp., San Jose, USA). At least 20 nuclei were analysed for 8 of the samples. The remaining sample had fewer cells suitable for analysis. Cells in adjacent normal epithelium were counted for 2 samples.

Results

Gain of 3q and 8q is associated with increased TERC and C-MYC copy numbers

As validation of the arrayCGH results, FISH analysis was performed using probes at 3q (TERC) and at 8q (C-MYC), both located at regions showing frequent gains in CIN3≥5yrPHI. Of the samples selected for FISH, 2 had no chromosomal aberrations as detected by arrayCGH, 4 had a gain of 3q, 1 had a gain of 8q and 2 samples had a gain of both 3q and 8q. In 6 samples the results fully matched that of the arrayCGH analysis (see Supporting Table II, for examples see Supporting Figure II). Two cases (CIN3.2.2 and CIN3.2.17) showed a gain at 3q by FISH analysis, which was not detected by arrayCGH, although in both cases a breakpoint indicative of a gain was seen in the chromosomal profiles. However, the probability of this aberration was too low in order to pass the threshold for being called as gain in the CGHcall algorithm. In a single case (CIN3.2.6) a gain at 8q was detected in 3% of cells, which was not perceived with arrayCGH. No breakpoint was witnessed in the segmented arrayCGH data and, due to the low number of dysplastic cells with this type of aberration, is likely to remain undetected by arrayCGH. In 2 samples adjacent normal tissue was analysed as a reference, both of which showed no copy number gains at 3q or 8q.
Chromosomal profiles of CIN3 relate to duration of hrHPV infection

Supporting Table I: Smallest regions of overlap (SRO) in CIN3<5yrPHI, CIN3≥5yrPHI and SCC.

<table>
<thead>
<tr>
<th>Cytoband</th>
<th>Aberration type</th>
<th>SRO size (kb)</th>
<th>Total</th>
<th>CIN3&lt;5yr PHI</th>
<th>CIN3≥5yr PHI</th>
<th>SCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>3q28</td>
<td>Gain</td>
<td>360.591</td>
<td>58.5</td>
<td>27.3</td>
<td>62.5</td>
<td>83.3</td>
</tr>
<tr>
<td>3q28</td>
<td>Gain</td>
<td>701.071</td>
<td>56.1</td>
<td>27.3</td>
<td>62.5</td>
<td>83.3</td>
</tr>
<tr>
<td>3q28</td>
<td>Gain</td>
<td>636.458</td>
<td>56.1</td>
<td>27.3</td>
<td>62.5</td>
<td>83.3</td>
</tr>
<tr>
<td>3q28</td>
<td>Gain</td>
<td>202.239</td>
<td>53.7</td>
<td>27.3</td>
<td>58.3</td>
<td>83.3</td>
</tr>
<tr>
<td>3q29</td>
<td>Gain</td>
<td>1653.419</td>
<td>53.7</td>
<td>27.3</td>
<td>58.3</td>
<td>83.3</td>
</tr>
<tr>
<td>3q29</td>
<td>Gain</td>
<td>492.912</td>
<td>51.2</td>
<td>27.3</td>
<td>54.2</td>
<td>83.3</td>
</tr>
<tr>
<td>3q29</td>
<td>Gain</td>
<td>213.49</td>
<td>51.2</td>
<td>27.3</td>
<td>54.2</td>
<td>83.3</td>
</tr>
<tr>
<td>3q26.1</td>
<td>Gain</td>
<td>133.842</td>
<td>48.8</td>
<td>27.3</td>
<td>50</td>
<td>83.3</td>
</tr>
<tr>
<td>3q28</td>
<td>Gain</td>
<td>1215.402</td>
<td>51.2</td>
<td>18.2</td>
<td>58.3</td>
<td>83.3</td>
</tr>
<tr>
<td>3q29</td>
<td>Gain</td>
<td>17.463</td>
<td>51.2</td>
<td>18.2</td>
<td>58.3</td>
<td>83.3</td>
</tr>
<tr>
<td>3q26.1-q26.31</td>
<td>Gain</td>
<td>12105.176</td>
<td>48.8</td>
<td>18.2</td>
<td>54.2</td>
<td>83.3</td>
</tr>
<tr>
<td>3q26.31-q28</td>
<td>Gain</td>
<td>13470.117</td>
<td>48.8</td>
<td>18.2</td>
<td>54.2</td>
<td>83.3</td>
</tr>
<tr>
<td>3q21.3-q26.1</td>
<td>Gain</td>
<td>32006.8</td>
<td>46.3</td>
<td>18.2</td>
<td>50</td>
<td>83.3</td>
</tr>
<tr>
<td>3q26.31</td>
<td>Gain</td>
<td>133.234</td>
<td>46.3</td>
<td>18.2</td>
<td>50</td>
<td>83.3</td>
</tr>
<tr>
<td>1p36.31-p36.13</td>
<td>Gain</td>
<td>9989.711</td>
<td>43.9</td>
<td>18.2</td>
<td>50</td>
<td>66.7</td>
</tr>
<tr>
<td>3q21.3</td>
<td>Gain</td>
<td>39.565</td>
<td>43.9</td>
<td>18.2</td>
<td>50</td>
<td>66.7</td>
</tr>
<tr>
<td>3q13.31-q21.1</td>
<td>Gain</td>
<td>7535.955</td>
<td>43.9</td>
<td>18.2</td>
<td>45.8</td>
<td>83.3</td>
</tr>
<tr>
<td>3q21.1-q21.2</td>
<td>Gain</td>
<td>2281.656</td>
<td>43.9</td>
<td>18.2</td>
<td>45.8</td>
<td>83.3</td>
</tr>
<tr>
<td>3q21.3</td>
<td>Gain</td>
<td>1385.975</td>
<td>43.9</td>
<td>18.2</td>
<td>45.8</td>
<td>83.3</td>
</tr>
<tr>
<td>1p36.31</td>
<td>Gain</td>
<td>308.221</td>
<td>41.5</td>
<td>18.2</td>
<td>45.8</td>
<td>66.7</td>
</tr>
<tr>
<td>3q13.13-q13.31</td>
<td>Gain</td>
<td>5124.729</td>
<td>41.5</td>
<td>18.2</td>
<td>45.8</td>
<td>66.7</td>
</tr>
<tr>
<td>1p35.3-p35.2</td>
<td>Gain</td>
<td>1181.205</td>
<td>39</td>
<td>18.2</td>
<td>45.8</td>
<td>50</td>
</tr>
<tr>
<td>1p32.3</td>
<td>Gain</td>
<td>421.088</td>
<td>39</td>
<td>18.2</td>
<td>45.8</td>
<td>50</td>
</tr>
<tr>
<td>1q21.3</td>
<td>Gain</td>
<td>88.751</td>
<td>39</td>
<td>18.2</td>
<td>45.8</td>
<td>50</td>
</tr>
<tr>
<td>1q21.3-q22</td>
<td>Gain</td>
<td>1033.091</td>
<td>39</td>
<td>18.2</td>
<td>45.8</td>
<td>50</td>
</tr>
<tr>
<td>3q13.11-q13.13</td>
<td>Gain</td>
<td>7096.395</td>
<td>39</td>
<td>18.2</td>
<td>45.8</td>
<td>50</td>
</tr>
<tr>
<td>3q21.1</td>
<td>Gain</td>
<td>119.913</td>
<td>41.5</td>
<td>18.2</td>
<td>41.7</td>
<td>83.3</td>
</tr>
<tr>
<td>3q21.2-q21.3</td>
<td>Gain</td>
<td>2435.762</td>
<td>41.5</td>
<td>18.2</td>
<td>41.7</td>
<td>83.3</td>
</tr>
<tr>
<td>3q21.3</td>
<td>Gain</td>
<td>75.215</td>
<td>39</td>
<td>18.2</td>
<td>41.7</td>
<td>66.7</td>
</tr>
<tr>
<td>1q21.3</td>
<td>Gain</td>
<td>213.65</td>
<td>36.6</td>
<td>18.2</td>
<td>41.7</td>
<td>50</td>
</tr>
<tr>
<td>1q41</td>
<td>Gain</td>
<td>690.079</td>
<td>36.6</td>
<td>18.2</td>
<td>41.7</td>
<td>50</td>
</tr>
<tr>
<td>1q42.2</td>
<td>Gain</td>
<td>822.858</td>
<td>36.6</td>
<td>18.2</td>
<td>41.7</td>
<td>50</td>
</tr>
<tr>
<td>1q31.3</td>
<td>Gain</td>
<td>6.112</td>
<td>46.3</td>
<td>9.1</td>
<td>54.2</td>
<td>66.7</td>
</tr>
<tr>
<td>1p36.31</td>
<td>Gain</td>
<td>24.672</td>
<td>39</td>
<td>9.1</td>
<td>45.8</td>
<td>66.7</td>
</tr>
<tr>
<td>1q31.3</td>
<td>Gain</td>
<td>128.868</td>
<td>39</td>
<td>9.1</td>
<td>45.8</td>
<td>66.7</td>
</tr>
<tr>
<td>1p22.1</td>
<td>Gain</td>
<td>77.108</td>
<td>36.6</td>
<td>9.1</td>
<td>45.8</td>
<td>50</td>
</tr>
<tr>
<td>1p22.1</td>
<td>Gain</td>
<td>186.351</td>
<td>36.6</td>
<td>9.1</td>
<td>45.8</td>
<td>50</td>
</tr>
<tr>
<td>1q21.1-q21.2</td>
<td>Gain</td>
<td>4456.563</td>
<td>36.6</td>
<td>9.1</td>
<td>45.8</td>
<td>50</td>
</tr>
</tbody>
</table>
**Percentages of regions that are frequently aberrated (occurring in ≥25.0% of the samples in all these groups combined). Aberrations on chromosomes 1 and 3 are already witnessed relatively frequently in CIN3<5yrPHI while aberrations on chromosomes 2 and 20 seem to occur later. Furthermore, the incidence of these aberrations increases according to progression stage.**

<table>
<thead>
<tr>
<th>Region</th>
<th>Type</th>
<th>Value</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q22-q23.2</td>
<td>Gain</td>
<td>4942.259</td>
<td>36.6</td>
<td>9.1</td>
<td>45.8</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q23.3-q25.2</td>
<td>Gain</td>
<td>15559.065</td>
<td>36.6</td>
<td>9.1</td>
<td>45.8</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q31.3</td>
<td>Gain</td>
<td>264.88</td>
<td>36.6</td>
<td>9.1</td>
<td>45.8</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q44</td>
<td>Gain</td>
<td>184.778</td>
<td>36.6</td>
<td>9.1</td>
<td>45.8</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1p36.32-p36.31</td>
<td>Gain</td>
<td>2456.357</td>
<td>36.6</td>
<td>9.1</td>
<td>41.7</td>
<td>66.7</td>
<td></td>
</tr>
<tr>
<td>1q31.3</td>
<td>Gain</td>
<td>75.819</td>
<td>36.6</td>
<td>9.1</td>
<td>41.7</td>
<td>66.7</td>
<td></td>
</tr>
<tr>
<td>3q21.3</td>
<td>Gain</td>
<td>285.534</td>
<td>36.6</td>
<td>9.1</td>
<td>41.7</td>
<td>66.7</td>
<td></td>
</tr>
<tr>
<td>1p22.1</td>
<td>Gain</td>
<td>961.178</td>
<td>34.2</td>
<td>9.1</td>
<td>41.7</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q25.2-q25.3</td>
<td>Gain</td>
<td>5326.723</td>
<td>34.2</td>
<td>9.1</td>
<td>41.7</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q25.3-q31.3</td>
<td>Gain</td>
<td>16141.997</td>
<td>34.2</td>
<td>9.1</td>
<td>41.7</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q31.3-q32.3</td>
<td>Gain</td>
<td>13151.31</td>
<td>34.2</td>
<td>9.1</td>
<td>41.7</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q32.3-q41</td>
<td>Gain</td>
<td>2153.415</td>
<td>34.2</td>
<td>9.1</td>
<td>41.7</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q41</td>
<td>Gain</td>
<td>5093.672</td>
<td>34.2</td>
<td>9.1</td>
<td>41.7</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q41</td>
<td>Gain</td>
<td>42.148</td>
<td>34.2</td>
<td>9.1</td>
<td>41.7</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q41-q42.13</td>
<td>Gain</td>
<td>5138.361</td>
<td>34.2</td>
<td>9.1</td>
<td>41.7</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q42.13</td>
<td>Gain</td>
<td>1752.79</td>
<td>34.2</td>
<td>9.1</td>
<td>41.7</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q42.13-q42.2</td>
<td>Gain</td>
<td>3325.179</td>
<td>34.2</td>
<td>9.1</td>
<td>41.7</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q42.2-q44</td>
<td>Gain</td>
<td>11633.397</td>
<td>34.2</td>
<td>9.1</td>
<td>41.7</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q44</td>
<td>Gain</td>
<td>4331.032</td>
<td>34.2</td>
<td>9.1</td>
<td>41.7</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1p36.32</td>
<td>Gain</td>
<td>249.372</td>
<td>34.2</td>
<td>9.1</td>
<td>37.5</td>
<td>66.7</td>
<td></td>
</tr>
<tr>
<td>1q25.3</td>
<td>Gain</td>
<td>110.88</td>
<td>36.6</td>
<td>9.1</td>
<td>37.5</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q21.2</td>
<td>Gain</td>
<td>619.7</td>
<td>34.2</td>
<td>9.1</td>
<td>37.5</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q32.3</td>
<td>Gain</td>
<td>669.565</td>
<td>31.7</td>
<td>9.1</td>
<td>37.5</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q41</td>
<td>Gain</td>
<td>146.208</td>
<td>31.7</td>
<td>9.1</td>
<td>37.5</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q41</td>
<td>Gain</td>
<td>62.074</td>
<td>31.7</td>
<td>9.1</td>
<td>37.5</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q41</td>
<td>Gain</td>
<td>556.003</td>
<td>31.7</td>
<td>9.1</td>
<td>37.5</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q42.13</td>
<td>Gain</td>
<td>113.186</td>
<td>31.7</td>
<td>9.1</td>
<td>37.5</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1q42.13</td>
<td>Gain</td>
<td>723.9</td>
<td>31.7</td>
<td>9.1</td>
<td>37.5</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>2q13.12</td>
<td>Gain</td>
<td>725.258</td>
<td>29.3</td>
<td>9.1</td>
<td>29.2</td>
<td>66.7</td>
<td></td>
</tr>
<tr>
<td>2q13.12</td>
<td>Gain</td>
<td>80.153</td>
<td>26.8</td>
<td>9.1</td>
<td>25</td>
<td>66.7</td>
<td></td>
</tr>
<tr>
<td>1p36.33</td>
<td>Gain</td>
<td>195.414</td>
<td>31.7</td>
<td>0</td>
<td>41.7</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>1p36.33-p36.32</td>
<td>Gain</td>
<td>1691.368</td>
<td>31.7</td>
<td>0</td>
<td>37.5</td>
<td>66.7</td>
<td></td>
</tr>
<tr>
<td>1p36.32</td>
<td>Gain</td>
<td>47.554</td>
<td>31.7</td>
<td>0</td>
<td>37.5</td>
<td>66.7</td>
<td></td>
</tr>
<tr>
<td>2q37.3</td>
<td>Loss</td>
<td>1010.461</td>
<td>26.8</td>
<td>0</td>
<td>33.3</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>
Chromosomal profiles of CIN3 relate to duration of hrHPV infection

Supporting Table II: Summary of 3q and 8q gains as determined by FISH as validation of the arrayCGH results.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Gains in arrayCGH</th>
<th>Number of cells counted</th>
<th>3q</th>
<th>8q</th>
<th>3q, 8q&lt;sup&gt;a&lt;/sup&gt;</th>
<th>n</th>
<th>%</th>
<th>3q</th>
<th>3q range</th>
<th>8q</th>
<th>8q range</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIN 3.2.15</td>
<td>-</td>
<td>26</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0</td>
<td>0%</td>
<td>1.54</td>
<td>0-2</td>
<td>0.96</td>
<td>0-2</td>
</tr>
<tr>
<td>CIN 3.2.17</td>
<td>-</td>
<td>31</td>
<td>19%</td>
<td>0%</td>
<td>0%</td>
<td>6</td>
<td>19%</td>
<td>1.77</td>
<td>0-4</td>
<td>0.97</td>
<td>0-2</td>
</tr>
<tr>
<td>CIN 3.1.8</td>
<td>3q</td>
<td>28</td>
<td>32%</td>
<td>0%</td>
<td>0%</td>
<td>9</td>
<td>32%</td>
<td>1.75</td>
<td>0-3</td>
<td>0.93</td>
<td>0-2</td>
</tr>
<tr>
<td>CIN 3.1.9</td>
<td>3q</td>
<td>11</td>
<td>45%</td>
<td>0%</td>
<td>0%</td>
<td>5</td>
<td>45%</td>
<td>2.55</td>
<td>1-4</td>
<td>0.73</td>
<td>0-2</td>
</tr>
<tr>
<td>CIN 3.2.6</td>
<td>3q</td>
<td>29</td>
<td>38%</td>
<td>0%</td>
<td>0%</td>
<td>12</td>
<td>41%</td>
<td>2.31</td>
<td>1-4</td>
<td>0.97</td>
<td>0-3</td>
</tr>
<tr>
<td>CIN 3.2.16</td>
<td>3q</td>
<td>39</td>
<td>38%</td>
<td>0%</td>
<td>0%</td>
<td>15</td>
<td>38%</td>
<td>2.15</td>
<td>0-4</td>
<td>1.18</td>
<td>0-2</td>
</tr>
<tr>
<td>CIN 3.2.2</td>
<td>8q</td>
<td>25</td>
<td>32%</td>
<td>16%</td>
<td>16%</td>
<td>16</td>
<td>64%</td>
<td>2.44</td>
<td>1-4</td>
<td>2.08</td>
<td>0-5</td>
</tr>
<tr>
<td>CIN 3.2.9</td>
<td>3q, 8q</td>
<td>41</td>
<td>32%</td>
<td>27%</td>
<td>5%</td>
<td>26</td>
<td>63%</td>
<td>2.12</td>
<td>0-4</td>
<td>2.07</td>
<td>0-4</td>
</tr>
<tr>
<td>CIN 3.2.13</td>
<td>3q, 8q</td>
<td>23</td>
<td>30%</td>
<td>9%</td>
<td>9%</td>
<td>11</td>
<td>48%</td>
<td>2.39</td>
<td>0-5</td>
<td>1.57</td>
<td>0-4</td>
</tr>
<tr>
<td>Normal, 3.2.9</td>
<td></td>
<td>29</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0</td>
<td>0%</td>
<td>0.52</td>
<td>0-2</td>
<td>1.07</td>
<td>1-2</td>
</tr>
<tr>
<td>Normal, 3.2.16</td>
<td></td>
<td>28</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0</td>
<td>0%</td>
<td>1.18</td>
<td>0-2</td>
<td>1.11</td>
<td>0-2</td>
</tr>
</tbody>
</table>

<sup>a</sup> % of cells positive for >2 probe signals for both 3q and 8q.
Supporting Figure I: Importance score plot between CIN3 in cluster 1 versus CIN3 in cluster 2. For each chromosomal region the maximum pair wise symmetrised Kullback-Leibler divergence was determined. A higher score indicates a larger contribution of that specific aberration to the difference between the CIN3 in the two clusters. In the plot can be seen that chromosome 3 is one of the most striking differences between CIN3 in the different clusters.
Supporting Figure II: Representative photographs of FISH results. In red 3q (TERC) and green 8q (C-MYC). (A) CIN3.2.15 (no 3q/8q), (B) CIN3.1.8 (3q gain), (C) CIN3.2.6 (3q gain), (D) CIN3.2.9 (3q/8q gain), (E) adjacent normal epithelium of CIN3.2.9, (F) adjacent normal epithelium of CIN3.2.16.